

# Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

<b>Autor práce</b>	Jakub Hajič		
<b>Název práce</b>	Zodpovídání dotazů o obrázcích		
<b>Rok odevzdání</b>	2017		
<b>Studijní program</b>	Informatika	<b>Studijní obor</b>	Umělá inteligence
<b>Autor posudku</b>	Milan Straka	<b>Role</b>	Vedoucí
<b>Pracoviště</b>	Ústav formální a aplikované lingvistiky		

### **Text posudku:**

Diplomová práce se zabývá úlohou *odpovídání otázek o obrázcích* (Visual Question Answering). Vstupem této úlohy je obrázek a otázka v přirozeném jazyce (např. „Kolik je na stole pomerančů?“ či „Vidí osoba na obrázku jako ostříž?“) a cílem je nalézt odpověď taktéž v přirozeném jazyce. V posledních dvou letech přilákala tato úloha velkou pozornost – došlo k vytvoření několika datových sad, konaly se dvě celosvětové soutěže, a výsledky navržených řešení jsou překvapivě dobré. Pokroky v řešení mají potenciálně řadu aplikací – jednak pro osoby se zhoršeným zrakem (např. fotografie plechovky a otázka „Jaké je datum spotřeby?“) či v cizím prostředí (např. fotografie japonského plánu a otázka „Kde je nejbližší bankomat?“), a také pro vyhledávání ve videu.

První dvě kapitoly popisují úlohu a seznamují čtenáře s přístupy používanými k jejímu řešení. Kapitola třetí poskytuje přehled o existujících datových sadách. Čtvrtá kapitola podrobně popisuje řešení (tzv. MCB model), založené na hlubokých konvolučních a rekurentních neuronových sítích, které obsadilo první místo v celosvětové soutěži VQA Challenge v roce 2016.

Těžiště práce se nachází v následujících třech kapitolách. Kapitola pátá popisuje dvě navrhovaná vylepšení MCB modelu, a popisuje dvě implementace těchto vylepšení (jedna pomocí frameworku Caffé a ve frameworku TensorFlow). Obě navrhovaná vylepšení a implementace jsou samostatná práce autora. Výsledky navrhovaných úprav jsou vyhodnoceny v kapitole šesté a dokumentace vytvořených implementací se nachází v kapitole sedmé.

Práci považuji za velmi kvalitní. Autor implementoval v té době nejlepší známý algoritmus pro úlohu odpovídání otázek o obrázcích, a navrhl a vyhodnotil několik jeho vylepšení. To vyžaduje velké technické dovednosti (použité architektury jsou velmi komplikované, samotné trénování na GPU trvá týden), odborné znalosti (navržení modifikací pokročilé neuronové sítě) a velké množství práce. Samotné výsledky prvního navrhovaného vylepšení jsou smíšené – na validační sadě jsou lepší než původní metoda, na test-dev sadě o něco horší. I když výsledky druhého navrhovaného vylepšení nepřekonal originalní metodu, v soutěži VQA Challenge 2017, která proběhla po odevzdání diplomové práce, zvítězila metoda zahraničního týmu, která byla založená na velmi podobné myšlence, což dokládá odbornou fundovanost autora.

Práce je napsaná velmi dobrou angličtinou a obsahuje rešerši současného stavu poznání. Dosažené výsledky je možno spolehlivě reprodukovat, protože implementace jak ve frameworku Caffé tak ve frameworku TensorFlow jsou zveřejněny včetně použitých hyperparametrů.

Celkově hodnotím diplomovou práci jako velmi povedenou a prokazující schopnost samostatné výzkumné činnosti.

**Práci doporučuji k obhajobě.**

**Práci nenavrhují na zvláštní ocenění.**

*Pokud práci navrhujete na zvláštní ocenění (cena děkana apod.), prosím uveďte zde stručné zdůvodnění (vzniklé publikace, významnost tématu, inovativnost práce apod.).*

**Datum** 28. srpna 2017

**Podpis**